

Disseny d'una activitat pedagògica musical amb visió per computador

Jordi Sans Pena

Resum— En un nou intent d'apropar la tecnologia a la societat i, més específicament, a les aules, en aquest projecte s'ha dissenyat i creat, amb la col·laboració de l'Escola de Música de Sant Cugat, un prototipus d'activitat didàctica enfocada a l'aprenentatge musical. En aquesta primera iteració de l'activitat, tres usuaris interactuen amb el seu cos i els seus gestos per fer sonar una peça musical. Un d'ells interactuarà amb un pentagrama gegant per fer sonar les notes d'un piano, trepitjant la nota que vulgui fer sonar. Els altres dos usuaris interaccionaran amb un ritme preenregistrat, podent alterar el seu tempo i volum mitjançant moviments. Aquest projecte combina diverses tecnologies per dur a terme el seu propòsit i crear una experiència didàctica que serveixi com a primer cop d'ull a les possibilitats que ens ofereix barrejar el món tecnològic amb el món de l'ensenyament.

Paraules clau— Educació, música, aprenentatge, visió per computador, integració, tecnologia, interacció.

Abstract— In a new attempt to bring technology to society and, more specifically, to classrooms, a prototype of educational activity focused in musical learning has been designed and created, with the collaboration of Sant Cugat's music school. In this first iteration of the activity, three users interact with their bodies and their gestures to make a musical piece play. One of them will interact with a giant stave to make piano notes sound, stepping on the note he wants to play. The other two users will interact with a pre-recorded rhythm, being able to modify its tempo and volume through their movements. This project combines different technologies to achieve its purpose and create a learning experience that will serve as a first look at the possibilities offered when mixing the technological world with the world of education.

Index Terms— Education, music, learning, computer vision, integration, technology, interaction.



1 INTRODUCCIÓ

La tecnologia s'ha convertit, en els darrers anys, en una part vital de la nostra vida i forma part de quasi totes les activitats que realitzem. Ja sigui a la feina, a casa, per desplaçar-nos o per comunicar-nos, la tecnologia sempre hi és present, i les aules d'una escola no són una excepció. Ordinadors, projectors o pissarres intel·ligents, ja formen part de les escoles i la introducció de noves tecnologies a l'ensenyança no s'aturarà aquí. En aquest context, tenim l'oportunitat d'aportar el nostre gra de sorra a construir una educació més interactiva. En col·laboració amb l'escola de música de Sant Cugat, en aquest projecte es duu a terme el disseny i la implementació d'una activitat pedagògica mitjançant la utilització de les noves tecnologies. Aquesta activitat ha de servir als alumnes més joves

de l'escola a interactuar amb la música d'una manera diferent de com ho farien amb els materials convencionals de les aules.

En aquest article hi trobarem 6 seccions: La primera secció està composta d'una petita introducció al projecte i dels objectius d'aquest. En la segona secció s'exposa la metodologia feta servir i una visió molt general de l'estructura de l'activitat i de quins actors hi intervenen. En la tercera secció es parla de l'estat de l'art de les diferents tecnologies fetes servir per a compondre el nostre prototip.

En la quarta secció s'explica de forma més detallada el funcionament de l'activitat i la implementació de l'última versió del prototipus. En la cinquena secció es recull una valoració dels resultats obtinguts amb l'activitat i, per últim, en la sisena secció podem trobar les conclusions del treball, quines millores es podrien aplicar i com es podria ampliar l'activitat.

-
- E-mail de contacte: jordisanspena@gmail.com
 - Menció realitzada: Computació.
 - Treball tutoritzat per: Fernando Vilariño (Computació)
 - Curs 2015/16

1.1 Objectius

L'objectiu principal del projecte, com ja s'ha exposat anteriorment, és el de dissenyar i implementar una activitat pedagògica, relacionada amb la música, que faci servir les noves tecnologies per funcionar.

Aquesta aplicació tindrà com a objectius la detecció d'un pentagrama gegant, que serà col·locat a terra i farà de perifèric principal per interactuar amb l'aplicació; i la detecció de la posició dels usuaris respecte a aquest pentagrama. Un cop aquestes deteccions s'aconsegueixin, l'aplicació haurà de respondre als moviments dels usuaris i actuar de manera adequada.

En aquesta primera iteració de l'aplicació hi participaran 3 usuaris, un d'ells interactuarà directament amb el pentagrama i els altres dos ho faran mitjançant els seus moviments. L'usuari del pentagrama haurà de poder fer sonar diferents notes segons on estigui col·locat, mentre que els altres dos usuaris interactuaran amb una base rítmica que sonarà en bucle.

Per aconseguir això, el primer pas serà estudiar quines opcions tecnològiques tenim al nostre abast per poder crear un prototip funcional de l'aplicació que serveixi com a incentiu per a invertir recursos en la creació d'una versió més refinada i preparada per a les aules.

2 METODOLOGIA

2.1 Scrum

La metodologia feta servir al projecte està inspirada en la metodologia Scrum [1], una metodologia de tipus àgil, però amb algun canvi en les seves característiques. Una de les característiques de Scrum és que tracta el desenvolupament d'un projecte de forma incremental, en comptes de planificar la totalitat d'aquest. Per això, primer s'ha dividit l'aplicació en les diferents funcionalitats que ha d'incorporar i s'han implementat de forma separada, validant el correcte funcionament de cadascuna per separat, és a dir, s'han implementat de forma seqüencial. Això difereix de Scrum, ja que en aquesta metodologia les diferents fases del desenvolupament se superposen.

2.2 Estructura bàsica

Com ja s'ha dit, en aquesta primera iteració de l'aplicació són tres els usuaris que hi podran interactuar. Un d'ells es podrà col·locar sobre el pentagrama gegant per fer sonar notes, en aquest cas, d'un piano. L'aplicació, a part, reproduirà una base rítmica amb la qual els altres dos usuaris podran interactuar. Un d'ells modificarà el volum al qual sona, modificant també el volum de les notes de l'usuari sobre el pentagrama, i l'altre modificarà el tempo d'aquesta base rítmica. Això ho aconseguiran mitjançant el moviment dels seus braços. A la Figura 1 hi podem veure una esquematització de l'escena.

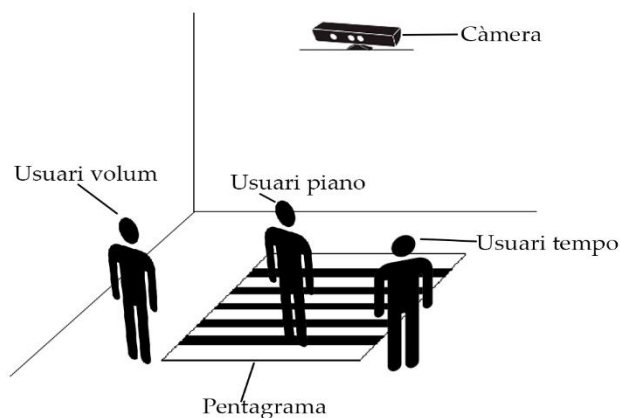


Fig. 1 Esquema bàsic de l'aplicació.

Com podem veure a partir de l'estructura bàsica de l'aplicació, hi ha diversos aspectes que el nostre programa ha de resoldre:

1. Detectar el pentagrama i les seves línies.
2. Reproduir els diferents sons involucrats en l'aplicació.
3. Detectar els usuaris i els seus moviments.
4. Respondre de forma adient a aquests moviments.

Per aconseguir resoldre cadascun d'aquests problemes, ens ajudarem de diferents tecnologies i software per tal de tenir una aplicació completa que ens permeti experimentar l'activitat docent que hem dissenyat.

3 ESTAT DE L'ART

El primer pas a l'hora d'implementar la nostra aplicació va ser escollir les diferents eines que ens han d'ajudar a resoldre els problemes exposats en l'apartat anterior.

El problema més important és el de decidir quin tipus de hardware farem servir per a la captura de la imatge, sobre la qual treballarem per detectar el pentagrama i els usuaris. Al nostre abast immediat disposàvem de diferents opcions entre les quals decidir, opcions tals com una càmera web o una Kinect [2,3], la càmera de Microsoft [4].

La principal diferència entre una càmera web normal i una Kinect és la quantitat d'informació que ens proporciona. Mentre que una càmera web només captura imatges, la Kinect, a part de també oferir imatge en color, està dotada d'un emissor de rajos infrarojos i d'un sensor de profunditat de rajos infrarojos que en conjunt fan possible la captura d'una imatge de profunditat, cosa que ens proporciona informació sobre la distància a la que es troben els diferents elements de l'escena respecte a la Kinect. Altres característiques que ofereix la Kinect sobre una càmera web estàndard són un conjunt de micròfons que ens permeten detectar la procedència dels sons enregistrats o, potser la característica més important, un software de seguiment de cossos.

Arribats a aquest punt, la decisió de decantar-se per fer

servir una Kinect a l'hora de desenvolupar la nostra aplicació sembla òbvia, no només per la informació extra que en podem extreure sinó que, a més, tenim a la nostra disposició el seu software de seguiment de cossos.

Tot i això, encara ens quedà una última decisió respecte al hardware que farem servir, quina versió de la Kinect utilitzarem? Tenim dues a la nostra disposició, i a la Taula 1 podem veure algunes de les seves diferències.

TAULA 1 Comparació entre la Kinect 1.0 i la Kinect 2.0

Característica	Kinect 1.0	Kinect 2.0
Càmera de color	640 x 480 píxels	1920 x 1080 píxels
Càmera de profunditat	320 x 240 píxels	512 x 424 píxels
Distància màxima de profunditat	4.5 m	4.5 m
Distància mínima de profunditat	40 cm	50 cm
Cossos complets detectats	2	6
Articulacions detectades als cossos	20	26

Com podem veure en aquesta petita comparativa, l'única opció vàlida per a la nostra aplicació és la versió 2.0 de la Kinect, ja que amb la primera versió només tenim informació sobre dues persones a l'escena i en necessitem tres com a mínim.

Ara que ja tenim resolt el problema de com capturar la nostra imatge, hem de decidir quines eines farem servir per tractar-la; i la decisió en aquest cas no és gaire difícil. OpenCv [5] és la millor aposta quant a llibreries de visió per computador, és programari lliure, té una quantitat de documentació molt gran i és una llibreria que s'utilitza mundialment. Amb OpenCv aconseguirem la detecció de les línies del pentagrama i també ens ajudarà a decidir les notes que hem de fer sonar quan l'usuari el trepitgi.

Existeixen diferents tècniques a l'hora de detectar línies en una imatge com, per exemple, la transformada de Hough [6,7], l'algorisme LSD [8] o algorismes basats en la intel·ligència d'eixam [9]. Dues d'aquestes opcions es troben implementades en les últimes versions d'OpenCv, la transformada de Hough i LSD, però són algorismes molt genèrics que busquen qualsevol línia en una imatge i nosaltres tenim un problema molt més específic que pot ser resolt d'una forma més senzilla, com veurem més endavant.

4 DESENVOLUPAMENT

En aquesta secció s'exposarà com s'han adreçat els diferents problemes o peces que formen part de la nostra aplicació i com s'han combinat per obtenir un prototip funcional.

Primer de tot, definim una mica millor quin serà el fil d'execució que la nostra aplicació ha de seguir per tal d'identificar totes les funcionalitats que el programa ha d'implementar. A la Figura 2 podem veure un diagrama de com el programa s'executarà normalment.

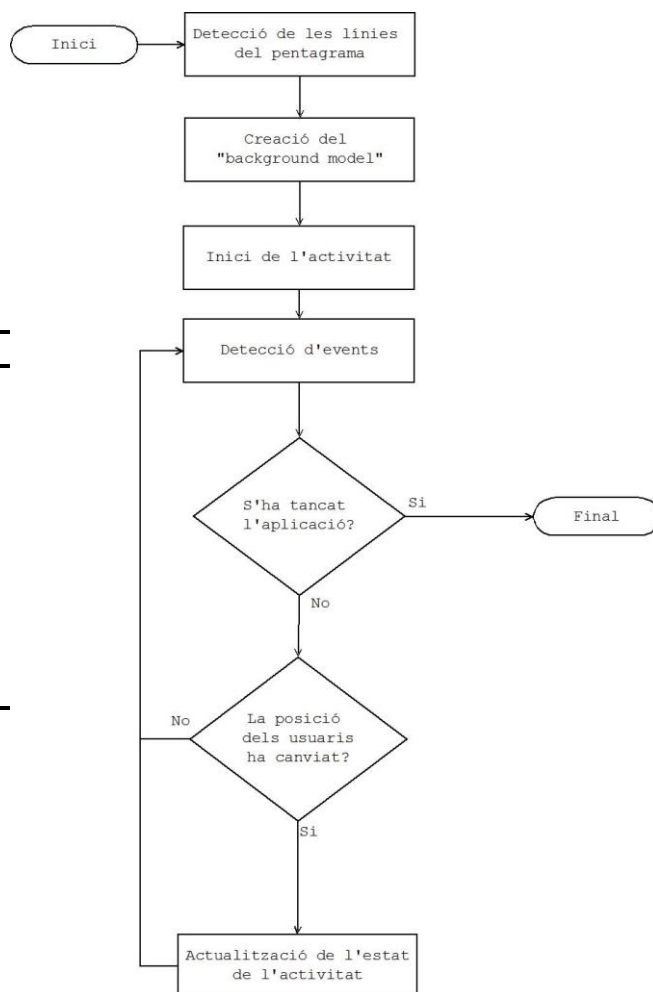


Fig. 2 Execució general del programa

Com podem veure en aquest diagrama de flux, el primer pas per al correcte funcionament del nostre programa és la detecció de les línies del pentagrama, ja que sense ell l'activitat perd el sentit. Per fer-nos una idea de l'escala amb què treballarem, a la Figura 3 podem veure el pentagrama real que s'ha fet servir per aquest prototip.



Fig. 3 Pentagrama

4.1 Detecció del pentagrama

Per a la detecció del pentagrama, s'han creat uns marcadors, amb el mòdul ArUcoMarkers [10,11,12] d'OpenCv, per tal de definir la zona on haurem de cercar el pentagrama. Gràcies a aquests marcadors, podem realitzar una cerca del pentagrama més ràpida i robusta, ja que hem delimitat l'espai de cerca a una zona coneguda. A la Figura 4 podem veure com queda configurada l'escena amb els marcadors afegits i detectats. La detecció d'aquests marcadors és molt robusta i ràpida, cosa que ens permet col·locar-los en qualsevol posició o angle respecte la càmera i que siguin detectats de forma eficaç.



Fig. 4 Detecció dels marcadors

Perquè la col·locació d'aquests marcadors sigui trivial per al docent que organitzi l'activitat, tots els marcadors involucrats són iguals i la seva orientació no és rellevant. L'única condició és que els marcadors formin un rectangle, que no cal que sigui perfecte, al voltant del pentagrama, col·locant un marcador a cada cantonada del pentagrama ja és suficient per al funcionament del programa.

Un cop els marcadors s'han detectat, el programa els ordena i crea la zona de joc, tal com veiem a la Figura 5, on l'usuari s'haurà de col·locar quan vulgui fer sonar les notes del piano.



Fig. 5 Detecció del pentagrama

El següent pas és trobar les línies del pentagrama dins aquesta zona que hem creat. El primer que farà el programa serà trobar els dos costats llargs del pentagrama, per tal de separar-lo del terra. Per a aquest propòsit pre-

pararem la imatge per treballar-hi, transformarem la imatge a escala de grisos i hi aplicarem un "threshold" adaptatiu.

Amb la imatge preparada, hi aplicarem una erosió amb un element horitzontal suficientment gran com per eliminar tots els elements de la imatge que no formin part del pentagrama. Ara ja tenim els recursos necessaris per buscar les nostres línies.

Crearem dues rectes que travessaran el pentagrama de forma perpendicular al costat llarg d'aquest i recorrerem aquestes rectes píxel a píxel per trobar els canvis de color, és a dir, canvis de línia. Aplicarem aquest procediment dues vegades, la primera sobre la imatge erosionada per tal de trobar on comença i on acaba el pentagrama i no tenir en compte el terra; la segona vegada l'aplicarem sobre la imatge sense erosionar, però les nostres rectes ara tindran els seus extrems en els límits del pentagrama que hem trobat en la iteració anterior. D'aquesta forma podrem detectar de forma precisa les línies del nostre pentagrama, tal com podem veure a la Figura 6. A la primera secció de l'apèndix podem veure el procés pas a pas.



Fig. 6 Línies del pentagrama detectades

4.2 Creació d'un "background model"

Un cop detectat el pentagrama i abans de començar l'activitat, encara ens queda una cosa més a fer.

A l'hora de detectar quines notes han de sonar respecte a la posició de l'usuari combinarem dues tècniques, una involucra el seguiment de la posició dels peus de l'usuari i l'altre serà un "background subtraction" [13]. Per dur a terme aquest "background subtraction", primer hem de crear el "background model" i això ho farem fent la mitjana dels 30 fotogrames següents a la detecció de les línies del pentagrama. Fent aquesta mitjana ens assegurem que tenim una imatge nítida com a model, i ens evitem que el model es vegi afectat pel soroll que pugui produir la Kinect.

4.3 Inici de l'activitat

Arribats a aquest punt, ja tenim tots els elements necessaris per començar l'activitat, però abans d'això definirem quines eines es fan servir per tractar els sons de l'aplicació i la imatge.

Per mostrar la imatge per pantalla i tractar els sons de les notes del pentagrama farem servir la llibreria SFML

[14], una llibreria multimèdia que incorpora diverses funcionalitats relacionades amb gràfics, àudio, xarxes, etc.

Farem servir SFML per carregar els fitxers on tenim guardat els sons de les notes i els podrem fer sonar cada vegada que l'usuari trepitgi el pentagrama. SFML fa servir diferents "threads" per fer sonar els diferents sons i, per tant, podrem fer sonar múltiples notes a la vegada si ho desitgem.

A més, farem servir SFML per crear una finestra per a la nostra aplicació on mostrarem la imatge que rebem de la Kinect i, també, farem servir el seu sistema d'esdeveniments per controlar quan apagar el programa.

D'altra banda, per fer sonar la base rítmica de l'aplicació farem servir la llibreria Bass [15] amb el seu mòdul BassFx. Aquesta llibreria, a diferència de SFML, és una llibreria especialitzada en el tractament del so i ens ofereix més possibilitats a l'hora de reproduir la nostra base que SFML.

La base es reproduirà en bucle al llarg de tota l'execució del programa, començant quan el "background model" s'hagi construït, i, com ja s'ha comentat, dos usuaris interactuaran amb ella modificant el volum i el tempo en temps real.

El control de volum és una característica que també podríem haver controlat amb SFML, però SFML no ens ofereix la possibilitat d'alterar el tempo de la base sense afectar al seu "pitch". Per aquest motiu BassFX és la llibreria triada per encarregar-se de la base rítmica, ja que ens ofereix control tant sobre el volum com sobre el tempo, minimitzant l'impacte d'aquests canvis sobre el "pitch".

4.4 Control dels usuaris

Ara ja podem començar l'activitat, i per això haurem de definir com l'aplicació interpreta els diferents moviments dels tres usuaris. La Kinect ens proporciona informació sobre els cossos que detecta a l'escena a cada fotografia, d'aquesta forma podrem comparar entre el fotograma anterior i l'actual els moviments dels nostres usuaris. Encara que l'activitat només tingui en compte tres usuaris, pot ser que dins l'escena aparegui alguna persona que no hi pertanyi i, per tant, haurem de filtrar quins cossos tenim en compte. Per fer això, tindrem en compte la informació 3D que rebem de la Kinect, informació que té les coordenades (0,0,0) al centre del sensor d'infrarojos. La coordenada X creix cap a l'esquerra del sensor, la coordenada Y cap amunt i la coordenada Z quan ens allunyem d'ell. Amb aquesta informació, definirem que les tres persones més properes a la Kinect, és a dir, amb els valors més petits de Z, seran els nostres usuaris; a més, amb la coordenada X podrem ordenar-los, definint l'usuari amb el valor de X més gran com l'usuari encarregat de manipular el tempo de la base rítmica, el segon usuari amb el valor de X més gran serà el que tocarà les notes i, per últim, l'usuari amb menor valor de X serà el que manipularà el volum.

Frame a frame enregistrem la posició d'aquests tres cossos i la compararem amb la del frame anterior, però no guardarem la mateixa informació pels tres usuaris.

4.4.1 Notes

Tal com hem definit l'activitat, l'usuari posicionat al centre de l'escena serà l'encarregat de fer sonar les notes trepitjant el pentagrama. Per tal que l'usuari es pugui moure pel pentagrama sense fer sonar notes que ell no desitgi que sonin, la nostra aplicació només es fixarà en la posició del seu peu dret, i serà aquesta informació la que anirem guardant frame a frame per comparar.

Per tal que soni una nota, la diferència en la coordenada Z del peu dret de l'usuari del frame actual i la del frame anterior no pot ser més gran d'un cert valor, que en aquesta versió és 0,0015 (expressat en metres). Aquesta mesura ens permet fer sonar notes només quan l'usuari trepitja en un cert punt del pentagrama i no durant tot el trajecte entre un punt i un altre. La posició que rebem de la Kinect pot variar lleument entre frame i frame encara que l'usuari no es mogui, per tant, no podem posar aquest "threshold" a 0.

Per evitar que una nota soni repetidament mentre que l'usuari està quiet, no pot sonar la mateixa nota dues vegades seguides. Amb una excepció, si l'usuari mou el peu en la coordenada X més de 0,05 metres, se li permetrà fer sonar la mateixa nota.

Aquestes regles, però, només ens defineixen quan podem fer sonar una nota, encara necessitem decidir quina nota hem de fer sonar. Per poder decidir, utilitzarem un "background subtraction" fent servir el model que hem creat a l'inici de l'aplicació i el frame actual. Podem veure el resultat d'aquest procediment a la Figura 7.



Fig. 7 Exemple del background subtraction

Per tal d'aprofitar la informació que ens proporciona aquest "background subtraction", crearem una finestra de 30 x 30 píxels al voltant de la posició 2D, dins la imatge, del peu dret i farem la mitjana de tots els píxels en blanc. Aquest càlcul ens proporcionarà la posició aproximada de la punta del peu de l'usuari, tal com podem veure a la Figura 8, on tots els punts vermells són proporcionats per

la Kinect i el punt groc és el calculat per l'aplicació.

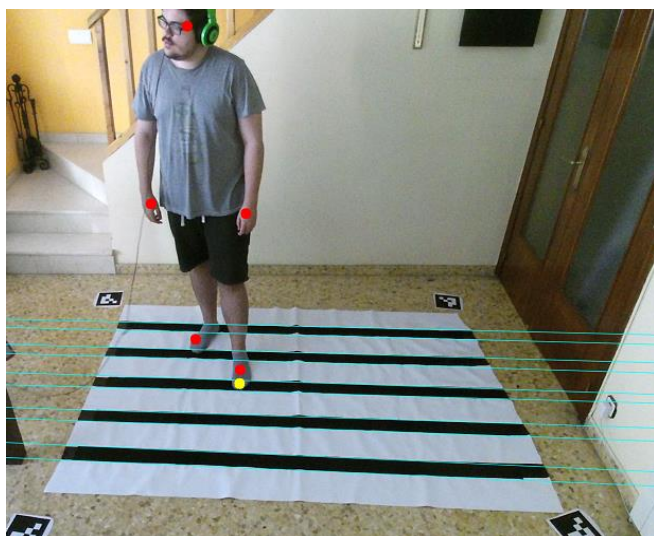


Fig. 8 Estimació de la posició del peu

Tenint la informació d'on són les línies del pentagrama i on és aproximadament el peu de l'usuari, ja podem decidir quina és la nota que ha de sonar. Ens ajudarem de (1), una equació que ens permet saber en quin costat d'una recta, definida pels punts A i B, es troba un punt P. Aquesta equació té en compte la direcció de les línies, per això ens hem d'assegurar de definir les línies del pentagrama d'esquerra a dreta.

$$\text{Side}(P) = (Bx - Ax) * (Py - Ay) - (By - Ay) * (Px - Ax) \quad (1)$$

Aquesta equació retorna valors positius quan el punt es troba a la dreta de la recta i valors negatius quan es troba a l'esquerra. De forma iterativa, comprovarem la posició del peu amb totes les rectes que tenim definides, començant per la recta més llunyana a la càmera fins a la més propera. Quan trobem una recta que en aplicar-li l'equació retorna un valor negatiu, sabrem que el peu està col·locat entre aquesta recta i l'immediatament anterior; sabent així quina nota hem de fer sonar.

4.4.2 Volum i tempo

Els usuaris encarregats de controlar el volum i el tempo de la base rítmica interactuen, en aquest primer prototip, de la mateixa forma amb l'aplicació. En aquest cas farem un seguiment de les mans dels usuaris, a diferència de l'usuari del pentagrama, per tal de decidir quan haurem d'augmentar el volum o el tempo i quan haurem de disminuir-los.

En aquest cas ens fixarem en les coordenades Y de les mans dels usuaris i comptabilitzarem les seves variacions al llarg de l'activitat per modificar les dues variables. Establirem dos límits en l'eix Y entre els quals els usuaris podran afectar a la base rítmica. El límit inferior serà la posició Y de la cintura i el límit superior serà la posició Y del cap, d'aquesta forma tenim un rang de moviment que s'adapta a la morfologia de cada usuari. Un cop tenim

aquests dos límits, calcularem la diferència absoluta entre ells i ho dividirem pel valor màxim que poden prendre les nostres variables, en el cas del volum és 100 per exemple. El que aconseguim amb aquesta divisió és l'increment unitari de cada variable en funció de la distància.

Calcularem les variacions en el volum/tempo dividint les diferències en l'eix Y de les mans dels usuaris entre el frame actual i l'anterior pels increments unitaris que hem calculat. Amb aquest càlcul augmentarem o disminuïrem el volum/tempo segons si els usuaris mouen els braços cap amunt o cap avall. El volum i el tempo tenen rangs de valors diferents entre ells, que podem veure a la taula 2. Volum 0 representa l'aplicació en silenci i volum 100 representa el màxim volum que el dispositiu que està executant l'aplicació pot suportar.

Els valors de tempo representen percentatges de velocitat en relació al tempo original, i com que aquests se sumen o resten al tempo original, un valor de -30 representa un tempo del 70% de la velocitat original de la peça, mentre que el valor màxim de +150 representa la peça original reproduïda 2,5 vegades més ràpida.

TAULA 2 Possibles valors de volum i tempo

Variable	Mínim	Màxim
Volum	0	100
Tempo	-30	+150

Una última condició que implementa aquesta versió de l'aplicació és que els dos braços de l'usuari han d'estar a la mateixa alçada, amb una certa tolerància, per tal de modificar el tempo/volum. Això implica que el moviment d'un sol braç no afectarà l'aplicació, i tampoc el moviment dels dos braços si aquests estan massa separats entre ells.

4.4.3 Integració

Un cop els diferents controladors definits treballen de forma conjunta, tenim una activitat completa en la qual els estudiants poden interactuar i crear música. Cada usuari pot operar sense tenir en compte la resta, però és quan els tres es compeneten quan es veuen les possibilitats d'una aplicació com aquesta.

5 RESULTATS

A causa de les característiques del treball, no tenim gaires mètriques quantitatives clares, però sí que podem observar com s'ha comportat el programa al llarg de les diferents execucions i com han reaccionat els diferents usuaris que l'han provat. La majoria de proves han sigut realitzades per mi o per coneguts al menjador de casa meua, però, dintre del procés final d'implementació, es va conduir una prova amb la col·laboració de l'Escola de Música. En aquesta prova hi van participar 4 professors i un grup de 10 alumnes, d'entre 10 i 12 anys; aquesta prova es va realitzar al Library Living Lab de Volpelleres [16], espai proporcionat per la UAB.

5.1 Detecció dels marcadors

La detecció dels marcadors per part de OpenCv és gairebé instantània, però, com moltes aplicacions de visió per computador, l'activitat ha de tenir una bona il·luminació.

L'únic detall a tenir en compte a l'hora de reproduir els resultats d'aquest projecte és que els marcadors han de ser capgirats horitzontalment respecte a la seva forma original. Això és produït per l'efecte mirall amb el qual la Kinect ens mostra la imatge.

5.2 Interacció amb el pentagrama

Des d'un punt de vista purament funcional, podem observar que com més propers a la càmera ens trobem, més acurada és la detecció de la nota que ha de sonar. La principal causa d'aquest fet és la perspectiva, les línies del pentagrama que es troben més a prop de la càmera apareixen més gruixudes que les més llunyanes.

Amb els espais o línies blanques del pentagrama aquest fet no afecta gaire, ja que són molt gruixudes, però amb les negres l'efecte és més visible i, encara que l'usuari tingui part del peu dins la línia, el càlcul de la posició del peu pot fer que soni la nota de l'espai en blanc següent.

A més, les posicions de les articulacions que rebem de la Kinect són menys fiables a distàncies més grans i, pel mateix usuari quiet, ens pot retornar posicions lleugerament diferents entre frames.

Per tal de poder visualitzar aquest fet de forma numèrica, s'ha dissenyat un petit experiment per testear la precisió de l'aplicació a l'hora de calcular les notes, tenint en compte només les línies negres. S'ha fet un recompte de quantes vegades l'aplicació calcula correctament la nota a tocar de cada 100 cops que es trepitja una línia. Aquest recompte s'ha realitzat 3 cops per cada línia negra diferent. Aquestes línies s'han numerat, essent la 1 la més propera a la Kinect i la 5 la més llunyana. A la Figura 9 podem veure una gràfica dels percentatges obtinguts i a la segona secció de l'apèndix hi podem trobar les dades mesurades.

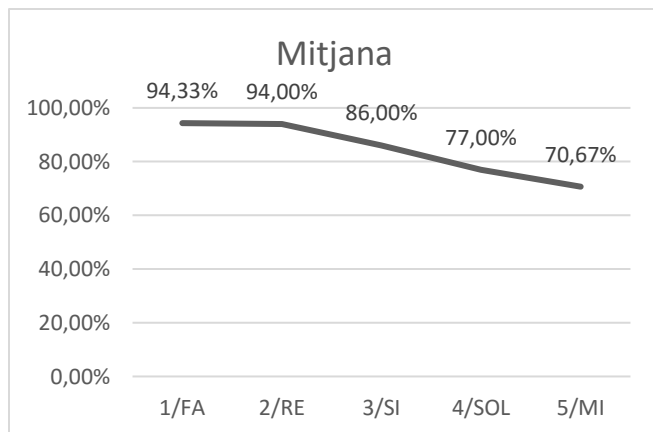


Fig. 9 Precisió del càlcul de notes

Tal com podem veure, a mesura que ens allunyem de la càmera, el percentatge d'encert a l'hora de calcular la nota adient disminueix.

La solució més simple seria fer més gruixudes totes les línies perquè la incertesa fos mínima.

D'altra banda, en observar les reaccions dels diferents usuaris també podem extreure certs resultats.

El control sobre les notes té diferents regles perquè l'usuari pugui fer sonar la nota que vulgui, i, amb usuaris més o menys madurs, aquestes regles no suposen cap problema, però per usuaris com els nens de l'Escola de Música que van provar el programa, aquestes regles no són del tot intuïtives. El fet que no es pugui repetir una nota només trepitjant, sinó que s'hagi de moure el peu de forma horitzontal, és el concepte que més conflicte va causar amb el que ells esperaven i s'haurà de buscar una forma més adient de solucionar el problema.

Un altre concepte rellevant és el fet que, degut a com l'aplicació decideix quan ha de sonar una nota, hi ha un cert retard entre que l'usuari trepitja una línia i la nota sona. Aquest retard és menor d'un segon, però és perceptible i s'haurà d'eradicar o, com a mínim, minimitzar en futures versions.

5.3 Controladors de volum i tempo

El control del volum/tempo és simple i robust, no importa com de ràpid l'usuari mogui els braços, l'increment o el decrement que s'aplicarà serà fidel als seus moviments. Però, tot i que funcionalment fa el desitjat, hi ha espai per a la millora. Els canvis són instantanis, és a dir, si en un fotograma tenim els braços per sobre del cap i en el següent per sota la cintura, passarem del volum/tempo màxim al mínim de forma instantània i, potser, una suavització d'aquesta transició seria interessant de forma conceptual.

6 CONCLUSIONS

L'objectiu principal d'aquest projecte era, en un període de temps reduït, ser capaçs de dissenyar i implementar una activitat pedagògica funcional que apropes la tecnologia a les aules. Podem dir que al final del projecte hem aconseguit satisfer aquest objectiu. És clar que perquè aquesta aplicació arribi a les aules s'haurà de millorar en certs aspectes i canviar en altres, però ens ofereix una primera visió del que es pot aconseguir fent servir la tecnologia que tenim a l'abast.

L'aplicació resultant combina diferents tecnologies i elements per aconseguir una experiència coherent. La Kinect ens proporciona una font d'informació molt valuosa, donant-nos informació RGB i de profunditat sobre la nostra escena. Amb OpenCv som capaçs de treballar sobre aquesta informació per tal de detectar un perifèric creat específicament per aquesta activitat i poder-lo fer servir, el pentagrama. SFML ens ofereix una interfície gràfica molt fàcil de crear i l'accés a diferents sons, mentre que BassFX ens permet aplicar modificacions més complexes sobre una base musical reproduint-se en temps real.

Per arribar a aquest punt, però, s'han hagut de treballar certs problemes que han anat apareixent al llarg del projecte. Una part molt important de l'aplicació és el pentagrama construït i els materials fets servir, en una primera versió les línies negres estaven fetes de cinta aïllant i això provocava reflexions amb la llum que dificultaven la detecció de les línies. Es va haver de fer servir un material de color negre mat per assegurar una millor robustesa.

Una altra part molt important és la correcta interpretació dels moviments dels usuaris perquè l'activitat respongui adequadament, i per garantir aquest fet s'han provat diferents aproximacions al problema. Per al control de volum/tempo, una primera aproximació no tenia un rang de moviments tancat sinó que s'adaptava a les posicions que les mans de l'usuari prenen durant l'execució. Aquest rang adaptatiu, en poder créixer en distància, implicava que es poguessin donar situacions en les quals no es representava de forma clara els canvis de volum/tempo respecte als moviments dels usuaris.

I és, un cop aquests problemes han sigut solucionats, quan tenim una aplicació completa sobre la qual seguir treballant i millorant per poder portar una nova experiència interactiva i divertida a les aules.

6.1 Treball futur

Com ja hem comentat, hi ha diverses formes en les quals podem millorar i expandir aquest projecte:

1. Millorar la construcció del pentagrama, fent les línies més gruixudes.
2. Afegir més usuaris, com un segon usuari que pugui interactuar amb el pentagrama. Creant així la possibilitat de tocar dues melodies diferents.
3. Canviar la forma amb la qual es controla el tempo perquè tingui més sentit conceptualment, per exemple, tenint en compte la quantitat de moviment que genera l'usuari.
4. Implementar algorismes més complexos per al control del tempo, per tenir millor qualitat d'àudio. Tenint en compte, això sí, el rendiment del programa.
5. Eliminar la reproducció per pantalla de l'escena, per evitar distraccions entre els usuaris i l'activitat.

Aquestes són només unes poques idees que es podrien explorar per tal de crear un programa més complet, però el pas més important per a crear una activitat digna de les aules seria treballar directament amb els docents. Dissenyar una activitat afegint la seva visió al projecte, menys tancada i matemàtica, i més orientada a com els alumnes esperen que l'aplicació reaccionï als seus moviments.

AGRAÏMENTS

Agrair, en primer lloc, a la meua parella, al meu millor amic i a la meua família la seva completa disponibilitat a ser els meus conillots d'índies per provar el programa a mesura que l'anava desenvolupant, la seva paciència en els moments de frustració i els seus ànims al llarg del camí. També agrair al meu tutor, Fernando Vilariño, la

seua ajuda en aquest projecte i la seva confiança en mi. Per últim, agrair a l'Escola de Música de Sant Cugat i als seus alumnes la seva col·laboració en aquesta experiència.

BIBLIOGRAFIA

- [1] Scrum Methodology, 2016; <http://scrummethodology.com/>
- [2] Kinect for Windows, 2016; <https://developer.microsoft.com/en-us/windows/kinect/>
- [3] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake, Real-Time Human Pose Recognition in Parts from Single Depth Images, 2011. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/BodyPartRecognition.pdf>
- [4] Microsoft, 2016; <https://www.microsoft.com/es-es/>
- [5] OpenCv, 2016; <http://opencv.org/>
- [6] Hough Transform, 2003; <http://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm>
- [7] Richard O. Duda, Peter E. Hart, Use of the Hough Transformation To Detect Lines and Curves in Pictures, 1972. <http://www.cse.unr.edu/~bebis/CS474/Handouts/HoughTransformPaper.pdf>
- [8] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, Gregory Randall, LSD: a Line Segment Detector, 2012. <http://dx.doi.org/10.5201/ipol.2012.gjmr-lsd>
- [9] Ulrich Kirchmaier, Simon Hawe, Klaus Diepold, A line detection and description algorithm based on Swarm Intelligence, 2010. <https://mediatum.ub.tum.de/doc/978673/978673.pdf>
- [10] ArUco Marker Detection, 2016; http://docs.opencv.org/master/d9/d6a/group__aruco.html
- [11] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, 2014. <http://www.uco.es/investiga/grupos/ava/sites/default/files/GarridoJurado2014.pdf>
- [12] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, R. Medina-Carnicer, Generation of fiducial marker dictionaries using mixed integer linear programming, 2015. https://www.researchgate.net/publication/282426080_Generation_of_fiducial_marker_dictionaries_using_Mixed_Integer_Linear_Programming
- [13] Massimo Piccardi, Background subtraction techniques: a review, 2004. https://profs.sci.univr.it/~cristanm/teaching/sar_files/lezione4/Piccardi.pdf
- [14] SFML, 2016; <http://www.sfm-dev.org/>
- [15] Bass audio library, 2016; <http://www.un4seen.com/>
- [16] Library Living Lab, 2016; <http://l3.cvc.uab.es/>

APÈNDIX

A1. DETECCIÓ DE LES LÍNIES DEL PENTAGRAMA

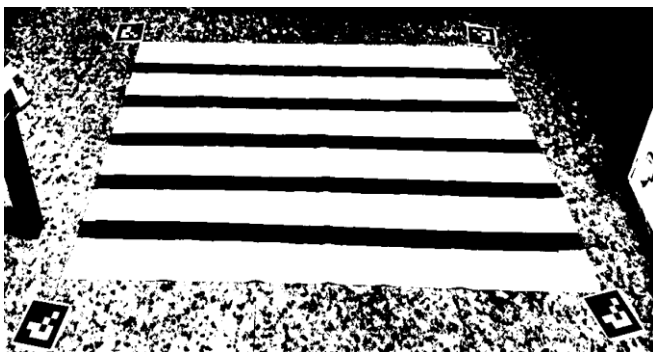
Detecció dels marcadors



Detecció del pentagrama



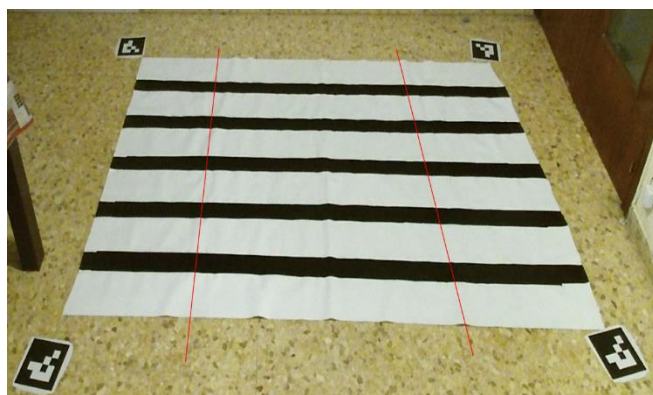
Threshold adaptatiu



Erosió de la imatge



Creació de les línies auxiliars



Detecció de les línies del pentagrama



A2. DADES MESURADES.

Primera Mesura:

Linea/Nota	OK	No OK	Percentatge d'encert
1/FA	96	4	96,00%
2/RE	93	7	93,00%
3/SI	81	19	81,00%
4/SOL	74	26	74,00%
5/MI	73	27	73,00%

Segona Mesura:

Linea/Nota	OK	No OK	Percentatge d'encert
1/FA	93	7	93,00%
2/RE	93	7	93,00%
3/SI	94	6	94,00%
4/SOL	76	24	76,00%
5/MI	67	33	67,00%

Tercera Mesura:

Linea/Nota	OK	No OK	Percentatge d'encert
1/FA	94	6	94,00%
2/RE	96	4	96,00%
3/SI	83	17	83,00%
4/SOL	81	19	81,00%
5/MI	72	28	72,00%